

# Determining Fundamental Properties of Matter Created in Ultrarelativistic Heavy-Ion Collisions

J. Novak, K. Novak, S. Pratt, and J. Vredevoogd

*Department of Physics and Astronomy and National Superconducting Cyclotron Laboratory,  
Michigan State University  
East Lansing, Michigan 48824, USA*

C.E. Coleman-Smith

*Department of Physics, Duke University  
Durham, North Carolina 27708, USA*

R.L. Wolpert

*Department of Statistical Science, Duke University  
Durham, North Carolina 27708, USA*

(Dated: March 27, 2013)

Posterior distributions for physical parameters describing relativistic heavy-ion collisions, such as the viscosity of the quark-gluon plasma, are extracted through a comparison of hydrodynamic-based transport models to experimental results from 100A GeV + 100A GeV Au+Au collisions at the Relativistic Heavy Ion Collider (RHIC). By simultaneously varying six parameters and by evaluating several classes of observables, we are able to explore the complex intertwined dependencies of observables to model parameters. We obtain a full multi-dimensional posterior distribution for the model output given a large set of experimental observations, the methods developed here provide a range of acceptable values for each parameter, and reveal correlations between them. The breadth of observables and the number of parameters considered here go far beyond previous studies in this field. The statistical tools, which are based upon Gaussian Process emulators, are tested in detail and should be extendable to larger data sets and a higher number of parameters.

PACS numbers: 25.75-q

Keywords: Heavy Ion Collisions, QGP, Bulk Properties, Gaussian Process, Model Data Comparisons, Emulator

## I. INTRODUCTION

Relativistic heavy ion collisions provide the means to study both the novel properties of the quark gluon plasma and the fascinating nature of how it is created and evolves. Unfortunately, experimental procedure is confined to measurements of the asymptotic momenta of the particles comprising the collision's debris. Addressing the fundamental questions concerning the properties of super-hadronic matter and the collision's evolution inherently depends on large-scale multistage transport models. Such models have improved significantly in recent years, and now typically combine viscous hydrodynamic treatments for the evolution of the semi-thermalized quark-gluon plasma ( $\sim 1\text{-}7$  fm/c), and microscopic hadronic simulations to describe the dissolution and breakup of the produced hadrons ( $\sim 7\text{-}20$  fm/c). For the first fm/c of the collision, when the system is too far from equilibrium for even a viscous hydrodynamic treatment, quantitative modeling carries large uncertainties. If the profile and flow of the matter being fed into the hydrodynamic treatment could be determined phenomenologically, it would be invaluable in understanding how QCD saturation phenomena affect the stopping of the matter, and in determining the mechanism and timescale for thermalization.

The data sets from the Relativistic Heavy Ion Collider (RHIC) and from the heavy ion programs at the Large

Hadron Collider (LHC) are immense. The heterogeneous nature of the data, along with the strong interdependence of disparate observables with respect to basic model parameters, makes any interpretation of the data challenging. The phenomenology of heavy ion collisions has progressed despite these difficulties, primarily by identifying the principal connections between model parameters and observables. For example, it is well understood that the shear viscosity of the quark-gluon plasma strongly affects the observed anisotropic flow coefficients. In an early analysis [1], the viscosity is adjusted until one finds a satisfactory fit with the anisotropic flow coefficient  $v_2$ . The shortcoming of such an approach is that several other unknown parameters, such as the spatial anisotropy of the initial state [26], also affect  $v_2$ . In turn, each of these parameters also affects numerous other observables. Similar approaches with more advanced models [2–10] have considered the variation of several parameters, and also the effects of such parameters on spectra. However, due to the numerical costs of running the models, these approaches have been unable to consider the simultaneous variation of more than two or three parameters, or to consider a wider range of experimental observables. These limitations compromise both the rigor and completeness of the effort.

Other fields of science face similar challenges, a notable case is the extraction of cosmological parameters

from observations of fluctuations of the cosmic microwave background [11, 12]. Here the parameters are some of the most fundamental in nature, such as the densities of dark matter and of dark energy. To overcome the limitations of running the model a large number of times, a surrogate model (a.k.a. an emulator) was developed to stand in for the true computer code. Rather than re-running the full cosmological evolution model during the exploration of the parameter space, one runs the full model at only  $\sim 100 - 200$  points in parameter space, carefully chosen to best fill the overall space. A surrogate model was constructed that effectively interpolated from the finite set of observations of the full model. The emulator was then substituted for the full model when exploring the parameter space. Similar ideas have been applied to the field of galaxy formation [13]. Here, we report first results for a large scale surrogate-model-based statistical analysis of heavy-ion collision data, a small scale application of these ideas was discussed in [14].

For this first effort, only a small subset of possible data will be considered, that coming from 100A GeV +100A GeV Au+Au collisions at RHIC. Spectra for pions, kaons and protons will be considered along with the elliptic flow observable  $v_2$  measured for pions, and femtoscopic source radii from two-pion correlations. The motivation for first considering soft observables is two-fold. First, they are the most sensitive to the model parameters related to the bulk properties of matter, and secondly, the dependencies are highly intertwined. During the last two years, the data set for relativistic heavy ion collisions has greatly expanded with the beam-energy scan at RHIC, and with the inaugural heavy-ion run at the LHC. The set will grow again in the next year as data is analyzed from Cu+Au and from U+U runs at RHIC. Ultimately, one may wish to incorporate other observables, such as dilepton emission, higher flow moments, or long-range correlations, once the theoretical treatments become more standardized and robust. The methods described here should scale well with increasingly large data sets, and incorporating additional observables into the analysis should be tractable.

On the theory side, numerous parameters factor into models of heavy-ion collisions. Several of these parameters are needed to describe the initial energy density and flow profiles the comprise the initial state of the hydrodynamic evolution. Other parameters describe the bulk properties of super-hadronic matter, such as the equation of state and viscosities. Still other parameters could describe out-of-equilibrium behavior such as chemical abundances of various quark species. For tractability only a half dozen parameters will be considered for this study. Four of the parameters describe the initial state for the hydrodynamic module, and two describe the shear viscosity and its energy dependence above the transition temperature. The equation of state from lattice calculations [15, 16] will be assumed to be correct. In a future study, that too will be parameterized to learn to what extent the equation of state is constrained ex-

perimentally. Hadronization will be assumed to produce a chemically-equilibrated hadronic gas when the energy density reaches 170 MeV. In the future, this assumption will also be relaxed and the away-from-equilibrium properties of these hadrons will be parameterized. Additionally, one should expect a non-negligible bulk viscosity in the transition region [17, 18]. However, due to some numerical instabilities with handling bulk viscosity, which will be set to zero for this study. An advantage of surrogate model techniques is that they scale well with an increasing number of parameters, and the efficiency of the methods are not greatly diminished if several parameters have only marginal impact. We expect these methods to continue to work even if we triple the number of parameters.

Details of the model and data used for the analysis are provided in the next two sections. The theory of the model emulator is described in Sec. IV, with a test of the emulator against a mock data set in the subsequent section. Results from an analysis of the real data set are given in Sec. VII, while a summary and outlook comprise the final section.

## II. MODELING THE EVOLUTION AT RHIC

For this study, four elements are involved in the modeling.

1. The pre-thermal, or stopping stage. Rather than dynamically solving for the evolution during this stage, we apply a parameterized description of the stress-energy tensor describing the state of the collision at a time of 0.8 fm/c. Although sophisticated models of the initial state do exist, e.g. [19–23], the large uncertainties and the lack of theoretical consensus dissuades one from picking any individual model.
2. The hydrodynamic stage lasts from 0.8 fm/c until the system falls below a hadronization temperature of 170 MeV. Viscous hydrodynamics is justified for a strongly interacting system that is not too far from equilibrium, and is especially convenient for a system undergoing a transition in degrees of freedom, because the equations can be applicable even when there are no well-defined quasi-particles.
3. Once the density has fallen to the point that the evolution can be modeled as binary collisions of hadrons, we switch to a microscopic hadronic simulation, or cascade. The cascade is able to handle the loss of equilibrium between species, e.g., the protons and pions moving with different average velocities or having different kinetic temperatures. The cascade also handles disassociation seamlessly.
4. Particles are correlated at small relative momentum due to interactions in the asymptotic state. Assuming that interactions with third bodies are

randomizing, one can calculate two-particle correlations given the source function, which describes the relative distance between two particles of similar velocities. Taking the source function from the information about last collisions in the cascade, and convoluting with the square of the known outgoing relative two-particle wave functions, we calculate correlations, and from the correlations calculate effective Gaussian source radii which can be compared to those extracted from experimentally measured correlations functions.

### A. Parameterizing the initial state

Rather than applying one of the competing models for the initial state, a parameterized form is used for the initial energy-density and flow profiles. Three parameters describe the initial energy-density profile and one

describes the flow profile. The first is a weight,  $f_{\text{wn}}$  between a wounded-nucleon profile and a saturation-based profile,

$$\epsilon(x, y) = f_{\text{wn}}\epsilon_{\text{wn}}(x, y) + (1 - f_{\text{wn}})\epsilon_{\text{sat}}(x, y). \quad (1)$$

The wounded-nucleon profile [24] and the saturation profiles are based on Glauber thickness functions which describe the projected areal densities of the incoming nuclei in a plane perpendicular to the beam axis,

$$T_{A,B}(x, y) = \int dz \rho_{A,B}(x, y, z), \quad (2)$$

where  $\rho_{A,B}$  are the baryon densities of the two nuclei given the impact parameters. The thickness functions have units of baryons per fm<sup>2</sup>, and the energy densities have the form,

$$\begin{aligned} \epsilon_{\text{wn}}(x, y) &= \frac{(dE/dy)_{pp}\sigma_{nn}}{2\sigma_{\text{sat}}} T_A(x, y) (1 - \exp(-T_B(x, y)\sigma_{\text{sat}})) , \\ &+ \frac{(dE/dy)_{pp}\sigma_{nn}}{2\sigma_{\text{sat}}} T_B(x, y) (1 - \exp(-T_A(x, y)\sigma_{\text{sat}})) , \end{aligned} \quad (3)$$

$$\epsilon_{\text{sat}}(x, y) = \frac{(dE/dy)_{pp}\sigma_{nn}}{\sigma_{\text{sat}}} T_{\min}(x, y) (1 - \exp(-T_{\max}(x, y)\sigma_{\text{sat}})) , \quad (4)$$

$$T_{\min} = \frac{2T_A T_B}{T_A + T_B}, \quad T_{\max} = (T_A + T_B)/2.$$

Here, the energy densities are per transverse area and per longitudinal rapidity, i.e., one would divide by  $\tau_0$  to get units per fm<sup>3</sup>. The three parameters are  $f_{\text{wn}}$ , the saturation cross section  $\sigma_{\text{sat}}$ , and the normalization  $(dE/dy)_{pp}$ . When two identical columns of nuclei collide,  $T_A = T_B$ , which leads to  $\epsilon_{\text{wn}} = \epsilon_{\text{sat}}$ . The quantity  $\sigma_{nn}$  is not an adjustable parameter, it is the known inelastic nucleon-nucleon cross section of 42 mb.

In the diffuse limit, where  $T_A, T_B \rightarrow 0$ , the energy density becomes  $(dE/dy)_{pp}T_A T_B \sigma_{nn}$ , which is known as the binary collision limit. If one considers two diffuse nuclei colliding randomly over a large area  $S$ , one finds the average energy per area in either expression to be

$$\langle dE/d\eta \rangle = \frac{\sigma_{nn}(dE/dy)_{pp}}{S} \int dx dy T_A(x, y) \int dx' dy' T_B(x', y') = \frac{AB\sigma_{nn}(dE/dy)_{pp}}{S}. \quad (5)$$

The parameter  $(dE/dy)_{pp}$  is the energy per unit rapidity of a  $pp$  collision at  $\tau_0$ . Although that number is measured in the asymptotic limit, it might be different for  $\tau_0 = 0.8$  fm/c. Thus, it is a parameter that is adjusted from 0.85 to 1.2 times the energy per rapidity of a  $pp$  collision of [25].

The parameter  $\sigma_{\text{sat}}$  controls the scale for changing the behavior of  $\epsilon_{\text{sat}}$  from the binary collision limit where  $\epsilon \sim T_A T_B$  to the saturated limit when  $\epsilon \sim T_{\min}$ . The change occurs for  $T_{\max} \approx 1/\sigma_{\text{sat}}$ . The parameter  $\sigma_{\text{sat}}$  also changes the wounded nucleon scaling form from that of binary collisions to the saturated limit where it is proportional to  $T_A + T_B$ .

The wounded nucleon and saturation expressions differ when  $T_a \neq T_b$ . For the case where  $\sigma_{\text{sat}}T_a \gg 1$  and  $\sigma_{\text{sat}}T_b \ll 1$ ,

$$\begin{aligned} \lim_{\substack{\sigma_{\text{sat}}T_a \gg 1 \\ \sigma_{\text{sat}}T_b \ll 1}} \epsilon_{\text{wn}} &= \frac{(dE/dy)_{pp}\sigma_{nn}}{\sigma_{\text{sat}}} T_a/2, \\ \lim_{\substack{\sigma_{\text{sat}}T_a \gg 1 \\ \sigma_{\text{sat}}T_b \ll 1}} \epsilon_{\text{sat}} &= \frac{(dE/dy)_{pp}\sigma_{nn}}{\sigma_{\text{sat}}} 2T_b. \end{aligned} \quad (6)$$

For a single nucleon,  $\sigma_{\text{sat}}T_b \ll 1$ , colliding onto a thick target,  $\sigma_{\text{sat}}T_a \gg 1$ , the energy density in the wounded nucleon expression continues to scale proportional to  $T_a$ . For example, colliding a single nucleon onto a target with  $\sigma_{\text{sat}}T_a = 10^6$  would give nearly 1000 times the multiplic-

ity for a collision with  $\sigma_{\text{sat}}T_a = 1000$ . In contrast, the saturation formula would give roughly the same energy density for both instances. It was shown in [26] that differences such as these significantly affect the initial elliptic anisotropy, and therefore significantly affect the measured elliptic flow. This can be understood by considering the collision of two equal mass nuclei with an impact parameter in the  $x$  direction. Along the  $x = 0$  line, both the wounded nucleon and saturation expressions give the same energy density. However, if one goes outward so that  $x$  becomes sufficiently large that one is at the edge of one nucleus, while being near the center of the other nucleus, the wounded nucleon formula gives a significantly higher energy density. This gives a relatively lower elliptic anisotropy for the wounded-nucleon model, and results in lower elliptic flow for the wounded-nucleon form than for the saturation form.

The fourth varied parameter describes the initial transverse flow, i.e., the collective flow at  $\tau_0 = 0.8 \text{ fm}/x$ . Initial flow has been found to significantly affect femtoscopic source sizes [27] and elliptic flow [28]. In [29, 30] it was shown that one can express the transverse flow as

$$\frac{T_{0i}}{T_{00}} = \frac{-\partial_i T_{00}}{2T_{00}}\tau, \quad (7)$$

given four conditions: (a) A traceless stress-energy tensor, (b) Lowest order in  $\tau$ , (c) Bjorken boost-invariance, (d) Anisotropy of the stress-energy tensor independent of  $x$  and  $y$ . The power of the parameterization is that in the high-energy limit one expects each of these conditions to be reasonably met. However, at finite energy and for higher orders in  $\tau$ , (7) can only serve as a guide to set a scale for the initial flow and cannot be trusted to better than a factor of two. For that reason, the initial flow is parameterized as a constant  $F_{\text{flow}}$  multiplied by the amount given in (7) for  $T_{0i}/T_{00}$ . The fraction  $F_{\text{flow}}$  was varied from 0.25 to 1.25.

For this first study, the initial energy density profiles are calculated from the average areal densities of the incoming nuclei, and are smooth, as if many events from the same impact parameter were averaged together. This is known to be fairly unrealistic, and the shortcoming will be addressed in the future.

## B. Hydrodynamic Module

Viscous hydrodynamics in an environment where there are no net conserved charges is based on local energy momentum conservation plus two assumptions. First, it is assumed that in the rest frame of the stress-energy tensor the effective pressure equals the equilibrated pressure,

$$\frac{T_{xx} + T_{yy} + T_{zz}}{3} = P(\epsilon), \quad (8)$$

i.e. the bulk viscosity is assumed to be zero. Second, it is assumed that the remainder of the stress-energy tensor is

sufficiently close its Navier Stokes value that its evolution can be described with Israel-Stewart equations of motion, which in the frame of the fluid becomes

$$\begin{aligned} \pi_{ij} &\equiv T_{ij} - \frac{1}{3}\delta_{ij}(T_{xx} + T_{yy} + T_{zz}), \\ \frac{d}{dt} \frac{\pi_{ij}}{\sigma(\epsilon)} &= -\frac{1}{\sigma(\epsilon)\tau_{\text{IS}}} \left( \pi_{ij} - \pi_{ij}^{(\text{NS})} \right), \\ \pi_{ij}^{NS} &= -\eta \left( \partial_i v_j + \partial_j v_i - \frac{2}{3}\delta_{ij} \nabla \cdot \mathbf{v} \right). \end{aligned} \quad (9)$$

The Israel-Stewart relaxation time was set to,  $\tau_{\text{IS}} = 3\eta/sT$ , which is the correct expression for an equilibrated gas of massless particles. Once these conditions are met, applying the local conservation of energy and momentum,

$$\partial_\mu T^{\mu\nu} = 0, \quad (10)$$

determines the evolution of the stress-energy tensor.

At high energy density the first assumption, that  $\sum_i T_{ii} = 3P(\epsilon)$ , can be met even if the system is far from chemical or kinetic equilibrium. For a gas of weakly interacting massless particles, or even for a region dominated by weakly interacting classical fields, the condition is met regardless of the configuration of either the particles or the fields. Once the fireball cools down near the transition region, and conformal invariance is lost, this assumption becomes questionable. The second assumption may be poorly met during the first 1-2 fm/c. However, the impact of changing the anisotropy of the stress-energy tensor at early times tends to be rather small [29].

The hydrodynamic module used here is built on an assumption of longitudinal boost-invariance which allows the calculations to become effectively two-dimensional before solving Israel-Stewart equations of motion. This approach has been applied by numerous research groups [31–34, 62]. The reduction of the dimensionality is justified to better than the five percent level [35]. The equation of state,  $P(\epsilon)$ , comes from lattice calculations of Wuppertal-Budapest group [15] for temperatures above the hadronization temperature, and use a hadron-gas equation of state at lower temperatures. The equation of state for temperatures just above the hadronization temperature is slightly modified from the lattice values to match the hadron gas value at the hadronization threshold.

For temperatures above 170 MeV/c, the viscosity to entropy density ratio was described with two parameters,

$$\frac{\eta}{s} = \frac{\eta}{s}|_{T_c} + \alpha \ln \left( \frac{T}{T_c} \right), \quad (11)$$

where  $T_c$  is assumed to be 170 MeV. The first parameter,  $\eta/s|_{T_c}$ , describes the viscosity just above the hadronization threshold, while the second parameter,  $\alpha$ , describes the temperature dependence.

The hydrodynamic/cascade interface temperature was set at a temperature of 170 MeV. Calculations were

also performed for a hadronization temperature of 155 MeV, but those calculations consistently over-predicted the flow, or equivalently, under-predicted the number of hadrons for a given amount of transverse energy. It is the author's intention to perform a detailed study of the sensitivity to the equation of state and the details of hadronization in a separate paper. A summary of model parameters is provided in Table. I.

### C. Hadronic Cascade

The hydrodynamic module was run until all elements cooled below 170 MeV. During the hydrodynamic evolution, the properties of the 170 MeV hypersurface were recorded. This included the boundaries, the flow velocity and the anisotropy of the stress-energy tensor. Hadrons were generated with a Monte Carlo procedure ensuring that all elements of the stress-energy tensor were continuous across the hyper-surface. The method [36] assumes that all species have a single time relaxation scale independent of their momentum. Other approaches have considered the effect of adding a momentum or species dependence to the relaxation time [32], but because this study considers only particles with low to moderate  $p_t$ , and because the particles interact a few more times in the cascade module, the details of the algorithmic choice are not expected to matter, as long as the stress-energy tensor is continuous across the boundary.

A list of particles as produced in the interface was then fed into the cascade on an event-by-event basis. For these studies, 4000 cascade events were produced for each impact parameter. The cascade code was inspired by the physics of the hadronic module of URQMD [37], but was significantly rewritten avto improve speed, and is labeled B3D [38]. Hadrons were assumed to collide through res-

onances with Breit-Wigner forms, plus a simple  $s$ -wave elastic cross section of 10 mb. The  $s$ -wave cross section was chosen independent of momentum and particle species. The resonances from the particle data book [39] with masses less than 2.2 GeV/ $c^2$  were all included. On average, particles collided less than twice after being generated from the hydrodynamic interface. Pions had fewer collisions on average, while protons had more. The collisions in the cascade mainly affected the spectra and  $v_2$  of protons. There are numerous ways to improve the cascade, such as more realistic cross sections, consistent time delays in scattering processes, mean-field effects, and Bose effects for pions. However, given the rather modest impact of the cascade at high energy, it is not expected that the observables would change significantly.

The B3D code runs approximately two orders of magnitude faster than URQMD for the calculations used here. This is mainly due to two improvements: better handling of the linked lists used to track collisions, and adding cyclic boundary conditions so that boost-invariance could be efficiently incorporated. The majority of the numerical expense of the calculations came from the cascade, and improving the speed allowed a greater number of points in parameter space to be explored.

The cascade ran until all collisions ceased. For each outgoing particle, the momentum, particle id and the space-time coordinates of the last interaction were recorded. Since the reaction plane is known, it is straightforward to calculate the azimuthal anisotropy factor  $v_2 = \langle \cos 2\phi \rangle$ . Spectra are efficiently calculated given that the cyclic boundary conditions make it possible to use all the particles when calculating the spectra at zero rapidity.

### D. Final-State Interactions

---

Two-particle correlations at small relative momentum provide femtoscopic information about the phase space distributions. This information is expressed through the Koonin formula [40, 41],

$$C(\mathbf{K} = (\mathbf{p}_1 + \mathbf{p}_2)/2, \mathbf{k} = (\mathbf{p}_1 - \mathbf{p}_2)/2) = \int d^3r S(\mathbf{K}, \mathbf{r}) |\phi_{\mathbf{k}}(\mathbf{r})|^2, \quad (12)$$

$$S(\mathbf{K}, \mathbf{r}) \equiv \frac{\int d^3r_1 d^3r_2 f(\mathbf{K}, \mathbf{r}_1) f(\mathbf{K}, \mathbf{r}_1) \delta(\mathbf{r} - (\mathbf{r}_1 - \mathbf{r}_2))}{\int d^3r_1 d^3r_2 f(\mathbf{K}, \mathbf{r}_1) f(\mathbf{K}, \mathbf{r}_1)}.$$

Here,  $\phi_{\mathbf{q}}(\mathbf{r})$  is the outgoing two-particle wavefunction,  $f(\mathbf{p}, \mathbf{r})$  is the phase space density in the asymptotic state, and  $S(\mathbf{K}, \mathbf{r})$  describes the chance that two particles with the same asymptotic momentum  $\mathbf{K}$  would be separated by  $\mathbf{r}$  should they not interact. Correlations provide the means to determine the coordinate-space information of  $S(\mathbf{K}, r)$  from the measured correlations,  $C(\mathbf{K}, \mathbf{k})$ . Through a fitting procedure, one can infer source radii which fit the shape of  $S(\mathbf{K}, \mathbf{r})$  with Gaussian radii, i.e.

---

$S(\mathbf{K}, \mathbf{r}) \sim \exp\{-x^2/2R_{\text{out}}^2 - y^2/2R_{\text{side}}^2 - z^2/2R_{\text{long}}^2\}$ , where the “outward” direction is transverse to the beam and parallel to  $\mathbf{K}$ , the “longitudinal” direction is along the beam axis and the “sideward” direction is perpendicular to the other two.

The source radii are typically extracted by experimental collaborations through fitting their measured correlations to expectations from Gaussian sources. Description of such analyses can be found in [41]. For the

parameter	description	range
$(dE/dy)_{pp}$	The initial energy per rapidity in the diffuse limit compared to measured value in $pp$ collision	0.85–1.2
$\sigma_{\text{sat}}$	This controls how saturation sets in as function of areal density of the target or projectile. In the wounded nucleon model it is assumed to be the free nucleon-nucleon cross section of 42 mb	30 mb–50 mb
$f_{wn}$	Determines the relative weight of the wounded-nucleon and saturation formulas for the initial energy density described in (3, 4)	0–1
$F_{\text{flow}}$	Describes the strength of the initial flow as a fraction of the amount described in (7)	0.25–1.25
$\eta/s _{T_c}$	Viscosity to entropy ratio for $T = 170$ MeV	0 – 0.5
$\alpha$	Temperature dependence of $\eta/s$ for temperatures above 170 MeV/c, i.e., $\eta/s = \eta/s _{T_c} + \alpha \ln(T/T_c)$	0 - 5

TABLE I. Summary of model parameters. Six model parameters were varied. The first four describe the initial state being fed into the hydrodynamic module, and the last two describe the viscosity and its energy dependence.

model calculations correlation functions were calculated by first sampling  $S(\mathbf{K}, \mathbf{r})$  then combining pairs of pions with similar momentum. Pions were divided into bins of 20 MeV/c width in transverse momentum and in  $15^\circ$  bins in azimuthal angle, before pairing. Utilizing boost invariance, all the pions could be longitudinally boosted to a frame where the rapidity was zero. The space-time points at which particle's had their last interaction had been recorded along with their asymptotic momentum during the running of the B3D module. This allowed a list of  $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$  to be constructed for each momentum bin. Correlation functions for each momentum bin were calculated by assuming a simplified wave function,  $|\phi_q(\mathbf{r})|^2 = 1 + \cos(2\mathbf{k} \cdot \mathbf{r})$ . Gaussian source radii were then found by searching for radii that best reproduce the correlation functions calculated by the model. Thus, rather than matching experimental and theoretical correlation functions, Gaussian radii were compared. The calculation of correlation functions and fitting was performed with the code base in CorAL [42].

### III. REDUCTION OF EXPERIMENTAL DATA FOR STATISTICAL ANALYSIS

The heavy ion data sets from RHIC and from the Pb+Pb experiments at the LHC represent some of the largest scientific data sets in existence. A principal motivation of this work is to develop a statistical analysis that can be extended to large heterogenous data sets. This would include data taken at multiple beam energies, with different target-projectile combinations and with different detectors. The recent beam-energy scan at RHIC and the inauguration of the LHC have increased the available data by more than an order of magnitude as compared to the Au+Au collisions at 100A GeV beams measured at RHIC. Additionally, analyzed measurements of Cu+Cu, Cu+Au and U+U from RHIC will soon be available. The data set from the one beam energy contains petabytes of information. For this first study, we confine our analysis to this one data set, Au+Au at 100A GeV + 100A GeV. We further confine the analysis to a subset of soft physics observables: spectra, elliptic anisotropy, and femtoscopic correlations. Only mid-rapidity observables were consid-

ered. These are the observables most connected to the bulk dynamics and to the bulk properties of matter, and are often referred to as “soft physics”. Several classes of observables are being ignored, e.g., jet quenching, long-range fluctuations and correlations, dilepton and direct-photon measurements, and heavy flavor. These observables are often labeled “rare probes” and their interpretation largely factorizes out of the analysis of the soft observables being considered here. For instance, although jet quenching depends on the energy density and bulk properties of the quark gluon plasma, the soft physics observables being considered here are not significantly affected by the mechanism for jet production. Further, the theory and phenomenology governing these other classes of observables often carry large uncertainties, not only in additional unknown parameters, but also in that they carry questions concerning the choice of approach. Given the way that the physics from these other classes of analyses factorize from the soft physics, and the lack of theoretical consensus, the prudent course of action seems to be to determine the bulk dynamics of the system using the soft physics observables. Once the evolution of the system is determined, with quantified uncertainties, one would have a validated basis from which to calculate other classes of observables, such as rare probes.

Within the set of soft-physics observables, this first analysis is restricted to a subset of the overall data. For spectra, we consider only pions, kaons and protons. It would be straight-forward to consider strange baryon spectra, but due to large systematic and statistical errors, they are unlikely to greatly affect the answer at the current time. Additionally, because theoretical treatments away from mid-rapidity remain in an immature stage, our analysis concerns only mid-rapidity observables. For angular anisotropies, we consider only  $v_2$  and ignore higher order anisotropies for  $n > 2$ ,

$$v_n \equiv \langle \cos(n\phi) \rangle, \quad (13)$$

where  $\phi$  is the angle of a particle relative to the reaction plane. Recent analyses of  $v_{n>2}$  suggest that the observables may even be more sensitive to the viscosity than  $v_2$  [43, 44]. However, theoretical questions remain about how to instantiate the event-by-event fluctuations which drive these higher-order harmonics. This analysis

only considers  $v_2$  for pions. Although  $v_2$  is measured for kaons and protons, in order to compare to data, theoretical treatments would have to run for tens of thousands of events for each impact parameter to get sufficient statistics for kaons and protons. This analysis used 4000 events per impact parameter. Finally, the femtoscopic analysis is confined to same-sign pions. Source sizes extracted from other analyses carry significantly more uncertainty. RHIC data is recorded according to centrality bins, e.g., top 5%, top 10%, 0-10% .... Bins are typically assigned according to some measure of overall multiplicity. For instance, the 20-30% bin corresponds to those events with multiplicities that are lower than the top 20% of events and higher than the lower 30% of events. The choice of bins varies between observables and between collaborations. Since the hydrodynamic treatment is questionable at low centrality, we decided to neglect the more peripheral collisions, i.e. those with centrality greater than 30%. Further, analysis was confined to two bins, 0-5% and 20-30%, due to the expectation that if those two bins were matched, any intermediate bin would also be matched. This reduced the numerical cost of performing the simulations.

It is our hope to extend future analyses to include more data. This would include data from the RHIC beam-energy scan, from the LHC, and from the Cu+Cu, Cu+Au and U+U at RHIC. Data from the LHC is straight-forward to incorporate because the same theoretical models can be used once one has added energy-dependence to the initial-state parameterization. The Cu+Au and U+U data require significantly rethinking the parameterization of the initial state, especially for Uranium due to large nuclear deformation. Extending the analysis to include data from the beam-energy scan would require significant changes to the model used here. At lower energies, one can no longer assume Bjorken boost-invariance and can no longer ignore the baryon excess. Although our present hydrodynamic code can work in three dimensions, significant theoretical work is required to develop a parameterization for the three-dimensional initial state at arbitrary beam energies.

### A. Initial Distillation of Observables

Experimental collaborations spend a lot of effort reducing the huge RHIC data set to a finite number of published plots representing a useful summary of some observable. For example, the PHENIX and STAR collaborations have produced plots of proton spectra for several centrality classes. Each plot might have a few dozen points. It is infeasible for our analysis to consider reproducing each of these data points. Instead, each observable was reduced to a few representative quantities. For a given species and centrality, spectra were reduced to two numbers. The first number is the yield, or integrated spectra, within a finite  $p_t$  (transverse momentum) range. The ranges were set so that they ignored the high

$p_t$  tail, which is strongly affected by jets and is outside the scope of the model. The second number would be the mean  $p_t$  within that range. The choice to use mean  $p_t$  was justified by doing a principal component analysis (PCA, see below) on the data points within a spectra divided by the yield. This ensured that only the shape of the spectra was coming into play.

To some extent such an analysis is inherently arbitrary, one needs to understand both theoretical and experimental systematic errors. By “theoretical systematic errors”, we refer to the fact that even if all the model parameters were correct, shortcomings of the model would preclude one from making comparing to data beyond a certain accuracy. Both classes of systematic error could have off-diagonal elements if one were to construct an  $N \times N$  error matrix, where  $N$  would be the number of data points in a specific spectra. The first principal component is sensitive to exactly how such a matrix would be generated. After trying several competing choices for systematic error, it was clear that the shape of the spectra was well represented by a single component. Further, that component was similar to the mean  $p_t$ . In essence, any two model runs that yield two spectra with the same  $p_t$  also yield indistinguishable spectra up to uncertainties. Thus, the spectra can be reasonably well summarized by the yield and mean  $p_t$ .

For the elliptic flow, the experimental information consisted of plots of  $v_2$  as a function of  $p_t$ . A PCA analysis showed that the  $p_t$ -weighted value for  $v_2$  effectively captured all the information within the set of model runs. Femtoscopic information was contained by plots from the STAR Collaboration of Gaussian radii ( $R_{out}$ ,  $R_{side}$  and  $R_{long}$ ) as a function of the transverse momentum. Simply averaging each radius over the several  $p_t$  bins was found to effectively encapsulate nearly all the variation of the femtoscopic radii throughout the model runs.

In this manner the various experimental results were thus reduced to those listed in Table. II. Each observable was also assigned an uncertainty. This uncertainty represented the accuracy within which a comparison of the theoretically determined value from a model run could be meaningfully compared to the corresponding experimental measurement. Of all the observables in Table. II only  $v_2$  has significant statistical error. Additionally, the  $v_2$  observable is known to be significantly affected by known shortcomings in the model, such as the lack of event-by-event fluctuations. Thus,  $v_2$  is assigned a larger percentage error than other observables.

### B. PCA Analysis of reduced observables

One could create model emulators for each of the observables listed in Table. II. However, one can further distill the data to a handful of principal components representing their most informative linear combinations. This serves to further reduce the complexity of the surrogate model. Let  $y_{exp,i}$  and  $\sigma_i$  be data points and uncertainties

observable	$p_t$ weighting	centrality	collaboration	uncertainty
$v_{2,\pi^+\pi^-}$	ave. over 11 $p_t$ bins from 160 MeV/c to 1 GeV/c	20-30%	STAR* [45]	12%
$R_{\text{out}}$	ave. over 4 $p_t$ bins from 150-500 MeV/c	0-5%	STAR [46]	6%
$R_{\text{side}}$	ave. over 4 $p_t$ bins from 150-500 MeV/c	0-5%	STAR [46]	6%
$R_{\text{long}}$	ave. over 4 $p_t$ bins from 150-500 MeV/c	0-5%	STAR [46]	6%
$R_{\text{out}}$	ave. over 4 $p_t$ bins from 150-500 MeV/c	20-30%	STAR [46]	6%
$R_{\text{side}}$	ave. over 4 $p_t$ bins from 150-500 MeV/c	20-30%	STAR [46]	6%
$R_{\text{long}}$	ave. over 4 $p_t$ bins from 150-500 MeV/c	20-30%	STAR [46]	6%
$\langle p_t \rangle_{\pi^+\pi^-}$	200 MeV/c < $p_t$ < 1.0 GeV/c	0-5%	PHENIX [47]	3%
$\langle p_t \rangle_{K+K^-}$	400 MeV/c < $p_t$ < 1.3 GeV/c	0-5%	PHENIX [47]	3%
$\langle p_t \rangle_{p\bar{p}}$	600 MeV/c < $p_t$ < 1.6 GeV/c	0-5%	PHENIX [47]	3%
$\langle p_t \rangle_{\pi^+\pi^-}$	200 MeV/c < $p_t$ < 1.0 GeV/c	20-30%	PHENIX [47]	3%
$\langle p_t \rangle_{K+K^-}$	400 MeV/c < $p_t$ < 1.3 GeV/c	20-30%	PHENIX [47]	3%
$\langle p_t \rangle_{p\bar{p}}$	600 MeV/c < $p_t$ < 1.6 GeV/c	20-30%	PHENIX [47]	3%
$\pi^+\pi^-$ yield	200 MeV/c < $p_t$ < 1.0 GeV/c	0-5%	PHENIX [47]	6%
$\pi^+\pi^-$ yield	200 MeV/c < $p_t$ < 1.0 GeV/c	20-30%	PHENIX [47]	6%

TABLE II. Observables used to compare models to data. \*To account for non-flow correlations, the value of  $v_2$  was reduced by 10% from the value reported in [45].

for the  $i = 1$  through  $N$  data points listed in Table. II. One then considers the corresponding quantities from the model run  $m$ ,  $y_{m,i}$  where  $m$  runs from 1 to the number of full model runs  $M$ . A useful first step is to scale the quantities by their net uncertainty,

$$\begin{aligned}\tilde{y}_{\text{exp},i} &= \frac{y_{\text{exp},i} - \langle y_i \rangle}{\sigma_i}, \\ \tilde{y}_{m,i} &= \frac{y_{m,i} - \langle y_i \rangle}{\sigma_i}, \\ \langle y_i \rangle &= \frac{1}{M} \sum_{m=1}^M y_{m,i}.\end{aligned}\quad (14)$$

The net uncertainties,  $\sigma_i$  are operationally defined as the uncertainty involved in comparing a model value to an experimental measurement. The measurements considered in this paper are limited by systematic rather than aleatoric errors, and we assume that errors are described by a normal distribution,

$$\mathcal{L}(\mathbf{x}) \sim \exp \left\{ - \sum_i \frac{\left( y_i^{(\text{exp})} - y_i^{(\text{mod})}(\mathbf{x}) \right)^2}{2\sigma_i^2} \right\}, \quad (15)$$

where  $y^{(\text{exp})}$  and  $y^{(\text{mod})}$  are the experimentally measured and model values respectively. Even if the model parameters are exact, the models also have limited accuracy due to shortcomings in the physics. Thus, the net uncertainty encapsulates both theoretical and experimental uncertainties, i.e., they can be considered to describe the inability of the model not only to describe the physics of the collision, but to also account for the inadequacy of the model to describe uncertainties in the experimental measurement and analysis.

One then studies the sample covariance of the model values amongst the  $M$  model runs.

$$S_{ij} = \frac{1}{M} \sum_{m=1}^M \tilde{y}_{m,i} \tilde{y}_{m,j} \quad (16)$$

The  $N$  eigenvalues of  $S$  are  $\lambda_i$ , and the normalized eigenvectors are  $\hat{e}_{i,j}$ . One can then consider new variables,  $z_{m,i}$  which are linear projections of the original  $\tilde{y}_{m,i}$  along the various directions defined by the eigenvectors,

$$z_{m,i} = \sum_j \hat{e}_{i,j} \tilde{y}_{m,j}. \quad (17)$$

With this procedure, the model values,  $\tilde{y}_{m,i}$ , are rotated into a basis where the values  $z_{m,i}$  have a diagonalized variance over the model runs,

$$\frac{1}{M} \sum_{m=1}^M z_{m,i} z_{m,j} = \lambda_i \delta_{ij}. \quad (18)$$

The values  $z_{m,i}$  are known as principal components. Since the values  $\tilde{y}$  were scaled by the uncertainties, the components  $\tilde{y}_i$  have uncertainties of unity, and after rotation the values  $z_i$  also have uncertainties of unity. Since the variance of  $z$  within the model runs is diagonal, one can state that those components for which  $\lambda_i \ll 1$  can be ignored because they do not assist in discriminating parameters. Further, the discriminating power is often dominated by the first few principle components, i.e., those with the largest  $\lambda_i$ .

To further justify our selection of principle components we show a plot of the normalized cumulative variance explained by the largest  $r$  components in Figure. 1, i.e

$$F(r) = \frac{\sum_{i=1}^r \lambda_i}{\sum_{i=1}^N \lambda_i}, \quad (19)$$

where we have sorted the eigenvalues into descending order. Examination of this figure clearly shows that the first four principle components are sufficient to explain almost all of the sampled variance.

Once the principal components,  $z_i$ , have been determined, one can invert the transformations to find  $y_i$  in terms of the  $z_i$ . The components which do not contribute



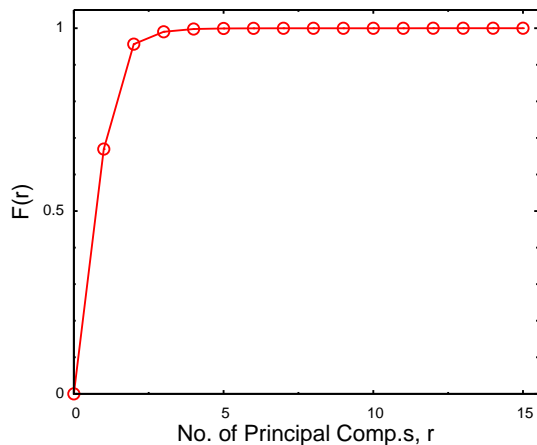


FIG. 1. The variance resolving power  $R(r)$  of the principal components, only the first few components are needed to explain almost all of the observed variance.

strongly to the total variance can be set to zero and the resulting  $y_i$ s will not be appreciably affected. In this particular case these are the components with  $\lambda \ll 1$ . Thus, the statistical analysis need only emulate those components with  $\lambda_i \gtrsim 1$ .

Given the 15 observables outlined in Table. II, one could construct an emulator for each observable. However, a PCA analysis of the 15 intermediate observables shows that not more than four of the principal components vary appreciably throughout the model runs. For these four components, the corresponding fluctuations,  $\langle \delta z_i^2 \rangle$  were of order unity or greater, while the remaining components fluctuated much less than unity. Thus, instead of tuning 15 emulators, only six principal components were considered (even though only four were truly needed). It is instructive to list the values of  $\lambda_i$  and the decomposition of the main components. This is shown in Table. III. The eigenvalues  $\lambda_i$  represent the resolving power of the various principal components.

To gain an understanding of the degree to which the parameter space is constrained by each measurement of  $z_i$ , one can consider the case where the prior distribution of parameters,  $x_\alpha$ , is distributed with unit variance according to a Gaussian distribution,  $\langle x_\alpha x_\beta \rangle = \delta_{\alpha\beta}$ . We further assume that each principal component  $z_j$  varies linearly with  $x$ . In that case, the gradient of each principal component,

$$(\nabla z_j)_\alpha \equiv \frac{\partial z_j}{\partial x_\alpha}, \quad (20)$$

form a set of orthogonal vectors because the covariance  $\langle z_i z_j \rangle = \lambda_i \delta_{ij}$  is diagonal,

$$\langle z_i z_j \rangle = (\nabla z_i)_\alpha (\nabla z_j)_\beta \langle x_\alpha x_\beta \rangle = \nabla z_i \cdot \nabla z_j = \lambda \delta_{ij} \quad (21)$$

Thus, if each component of  $z$  depends linearly on  $x$ , each principal component constrains a separate direction in parameter space. One can then understand the resolving

power by considering the simple case with one principal component and one parameter. Given a measurement  $z_{(\text{exp})}$  and assuming that the prior has unit variance and that  $z$  depends linearly with  $x$ ,

$$P(x) \sim e^{-x^2/2} \exp \left\{ - (mx - z_{(\text{exp})})^2 / 2 \right\}, \quad (22)$$

where  $dz/dx$  is the slope  $m$ , and from (21)  $m^2 = \lambda$ . Completing the squares in the argument of the exponential,

$$P(x) \sim e^{-(\lambda+1)(x-\mu^2)/2}, \quad (23)$$

where  $\mu$  is the prior mean for  $x$ . This shows that if the response is purely linear that each principal component reduces the width of the posterior relative to the prior by a factor  $1/\sqrt{1+\lambda_i}$ .

The first principal component in Table. III carries the bulk of the resolving power. Since  $\lambda_1 = 18.36$ , the linear considerations above suggest that a measurement of the first principal component should constrain the original parameter space by a factor of roughly  $1/\sqrt{19.36}$ . The second and third principal components also significantly narrow the parameter space. All together, one expects these measures to constrain the fit to on the order of 5% of the original six-dimensional parameter space at the “one-sigma” level. This estimate of the resolving power is based on an assumption that  $z$  varies linearly with  $x$ , but nonetheless provides a useful, although crude, expectation for how the our analysis might ultimately constrain the parameter space.

The first and second components dominantly consist of measures of the multiplicity and of the  $v_2$  observable. This is not surprising. It shows that the most important aspect of fitting data is to fit the multiplicity and elliptic flow. The third component has a large mixture of  $\langle p_t \rangle$  and interferometric observables. Thus, before performing the parameter space exploration, one expects that those parameters driving the multiplicity and elliptic flow will be the most significantly constrained.

#### IV. THEORY OF MODEL EMULATORS

Developing an understanding of a six-dimensional parameter space requires many millions of samplings of the model output. Running a complex code for each sampled point in parameter space is impractical. An alternative strategy has been to develop model emulators, codes which effectively interpolate from an initial sampling of runs through the space.

We construct a Gaussian Process emulator [48–51], which acts as a statistical model of our computer model. An emulator is constructed by conditioning a prior Gaussian Process on a finite set of observations of model output, taken at points dispersed throughout the parameter space. Once the emulator is trained it can rapidly give predictions for both model outputs and an attendant measure of uncertainty about these outputs at any point

observable \ $\lambda_i$	18.36	7.87	0.93	0.21	0.04	0.012
cent0to5_PHENIX_spectraPION_YIELD	0.43202	0.52170	0.21636	0.56290	0.06883	0.35417
cent0to5_PHENIX_spectraPION_MEANPT	0.10117	0.02647	0.37032	-0.08869	0.09235	-0.24640
cent0to5_PHENIX_spectraKAON_MEANPT	0.10770	0.03291	0.37755	-0.07459	0.06328	-0.26766
cent0to5_PHENIX_spectraPPBAR_MEANPT	0.04925	0.02192	0.16751	-0.02131	-0.05466	-0.19039
cent0to5_STAR_ROUT_PION	-0.01942	0.06908	-0.31734	-0.02968	0.72626	0.12886
cent0to5_STAR_RSIDE_PION	0.09148	0.09321	0.07972	0.05565	0.11943	-0.07137
cent0to5_STAR_RLONG_PION	0.08413	0.09520	-0.13599	0.37546	0.08521	-0.50343
cent20to30_PHENIX_spectraPION_YIELD	0.43743	0.49869	-0.32721	-0.56043	-0.26805	-0.01286
cent20to30_PHENIX_spectraPION_MEANPT	0.07549	0.03028	0.33981	-0.23142	0.28313	-0.06472
cent20to30_PHENIX_spectraKAON_MEANPT	0.08266	0.03721	0.34043	-0.23645	0.27941	-0.06785
cent20to30_PHENIX_spectraPPBAR_MEANPT	0.03791	0.02697	0.14297	-0.11517	0.03747	-0.09339
cent20to30_STAR_V2_PION_PTWEIGHT	-0.74299	0.65843	0.08846	-0.03607	-0.01531	-0.06192
cent20to30_STAR_ROUT_PION	0.02955	0.03296	-0.30420	-0.06375	0.43249	-0.09820
cent20to30_STAR_RSIDE_PION	0.08368	0.09367	-0.01379	-0.21381	0.08021	-0.06598
cent20to30_STAR_RLONG_PION	0.08974	0.08905	-0.24592	0.19088	-0.07458	-0.62873

TABLE III. The first six principle components. Since the variables were initially scaled by their uncertainties, the eigenvalues,  $\lambda_i$ , describe the resolving power of the components. Only the first  $\sim 4$  components are significant, i.e.,  $\lambda \gtrsim 1$ . The table also provides the decomposition of the principal components in terms of the 15 observables.

in the parameter space. This is a probability distribution for the model output at all points in parameter space and is by far the useful feature of Gaussian Process emulators. The most common interpolation schemes, such as interpolating polynomials, produce an estimate of the model output at a given location in the parameter space with no indication as to the extent that this value should be trusted. Furthermore, numerical implementations of Gaussian Process emulators are computationally efficient (producing output in microseconds rather than minutes), making it feasible to predict vast numbers of model outputs in a short period of time. This ability opens new doors for the analysis of computer codes which would otherwise require unacceptable amounts of time [52, 53].

We construct an emulator for a model by conditioning a Gaussian Process prior (see Figure. 2) on the training data [54–56]. A Gaussian Process is a stochastic process with the property that any finite set of samples drawn at different points of its domain will have a multivariate-normal (MVN) distribution. Samples drawn from a stochastic process will be functions indexed by a continuous variable (such as a position or time) as opposed to a collection of values as generated by, e.g., a normally-distributed random variable. A Gaussian Process is completely specified in terms of a mean and covariance, both of which can be functions of the indexing variable  $x$ . The covariance,  $c(x_1, x_2)$ , might be any positive-definite function of  $x_1$  and  $x_2$ . An example of unconditioned draws is shown in the left panel of Figure. 2 for the case where the covariance depends only on  $x_1 - x_2$  and is a power-exponential covariance function with unit length. The draws are smooth functions over the domain space, and if enough samples are drawn from the process the average of the resulting curves at each point would converge to zero.

A predictive distribution for the value of a computer model at new points in the design space can be obtained by conditioning this process on a set of training points

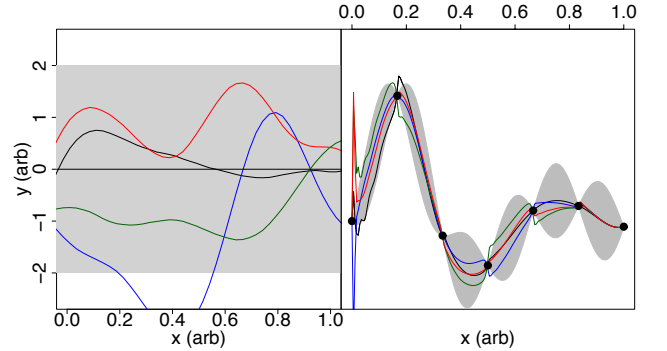


FIG. 2. Left panel: Unconditioned draws from a Gaussian Process  $GP(0, 1)$  with a mean of zero and constant unit variance. Right panel: draws from the same process after conditioning on 7 training points (black circles). The gray band in both panels is a pointwise 95% confidence interval. Note how the uncertainty in the right panel grows when away from the training points.

obtained from running the model. Conditioning forces samples drawn from the process to always pass through the training points. The resulting curves interpolate the training data, as shown in the right hand panel of Figure. 2. Repeated draws from the conditioned posterior distribution would on average follow the underlying curve with some variation, shown by the gray confidence regions. These confidence bubbles grow away from the training points, where the interpolation is least certain, and contract to zero at the training points where the interpolation is absolutely certain. The posterior distribution can be evaluated to give a mean and variance at any point in the parameter space. We may interpret the mean of the emulator as the predicted value at a point, the variance at this point gives an indication of how close the mean can be expected to be to the true value of the model.

To construct an emulator we need to fully specify our Gaussian Process (GP) by choosing a prior mean and a form for the covariance function. The model parameter space is taken to be  $p$ -dimensional. We model the prior mean by linear regression with some basis of functions

$\mathbf{h}(x)$ . In this analysis we use the trivial basis  $\mathbf{h}(x) = \{1\}$ . We specify a power exponential form for the covariance function with power  $\alpha \simeq 2$  to ensure smoothness of the GP draws ( $\alpha$  has to be in  $[1, 2]$  to ensure positive definiteness),

$$c(\mathbf{x}_i, \mathbf{x}_j) = \theta_0 \exp \left( - \sum_{k=1}^p \left\{ \frac{x_i^k - x_j^k}{\theta^k} \right\}^\alpha \right) + \delta_{ij} \theta_N, \quad \alpha \in [1, 2]. \quad (24)$$

Here,  $\theta_0$  is the overall variance, the  $\theta^k$  set characteristic length scales in each dimension in the parameter space and  $\theta_N$  is a small term, usually called a nugget, added to ensure numerical convergence or to model some measurement error in the code output. The shape of the covariance function sets how the correlations between pairs of outputs vary as the distance between them in the parameter space increases. The scales in the covariance function  $\theta^k$  are estimated from the data using maximum likelihood methods [56], in Figure. 3 we demonstrate their influence on an artificial data set. The linear regression model handles large scale trends of the model under study, and the Gaussian Process covariance structure captures the residual variations.

Given a set of  $n$  design points  $\mathcal{D} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  in a  $p$ -dimensional parameter space, and a set of  $n$  training values representing the model output at the design locations  $\mathbf{Y} = \{y_1, \dots, y_n\}$ , the posterior distribution defining our emulator is

$$\mathcal{P}(\mathbf{x}, \theta) \sim \text{GP} \left( \hat{m}(\mathbf{x}, \theta), \hat{\Sigma}(\mathbf{x}, \theta) \right), \quad (25)$$

for conditional mean  $\hat{m}$  and covariance  $\hat{\Sigma}$ .

$$\begin{aligned} \hat{m}(\mathbf{x}) &= \mathbf{h}(\mathbf{x})^T \hat{\beta} + \mathbf{k}^T(\mathbf{x}) \mathbf{C}^{-1} (\mathbf{Y} - \mathbf{H} \hat{\beta}), \\ \hat{\Sigma}(\mathbf{x}_i, \mathbf{x}_j) &= c(\mathbf{x}_i, \mathbf{x}_j) - \mathbf{k}^T(\mathbf{x}_i) \mathbf{C}^{-1} \mathbf{k}(\mathbf{x}_j) + \Gamma(\mathbf{x}_i, \mathbf{x}_j), \\ \mathbf{C}_{ij} &= c(\mathbf{x}_i, \mathbf{x}_j) \end{aligned} \quad (26)$$

$$\begin{aligned} \Gamma(\mathbf{x}_i, \mathbf{x}_j) &= (\mathbf{h}(\mathbf{x}_i)^T - \mathbf{k}^T(\mathbf{x}_i) \mathbf{C}^{-1} \mathbf{H})^T (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \\ &\quad (\mathbf{h}(\mathbf{x}_j)^T - \mathbf{k}^T(\mathbf{x}_j) \mathbf{C}^{-1} \mathbf{H}), \\ \mathbf{k}(\mathbf{x})^T &= (c(\mathbf{x}_1, \mathbf{x}), \dots, c(\mathbf{x}_n, \mathbf{x})). \end{aligned} \quad (27)$$

Where  $\hat{m}(\mathbf{x})$  is the posterior mean at  $\mathbf{x}$ ,  $\hat{\Sigma}(\mathbf{x}_i, \mathbf{x}_j)$  is the posterior covariance between points  $\mathbf{x}_i$  and  $\mathbf{x}_j$ ,  $\mathbf{C}$  is the  $n \times n$  covariance matrix of the design  $\mathcal{D}$ ,  $\hat{\beta}$  are the maximum-likelihood estimated regression coefficients,  $\mathbf{h}$  the basis of regression functions and  $\mathbf{H}$  the matrix of these functions evaluated at the training points.

The elements of the vector  $\mathbf{k}(\mathbf{x})$  are the covariance of an output at  $\mathbf{x}$  and each element of the training set. It is through this vector  $\mathbf{k}(\mathbf{x})$  that the emulator “feels out” how correlated an output at  $\mathbf{x}$  is with the training set and thus how similar the emulated mean should be to

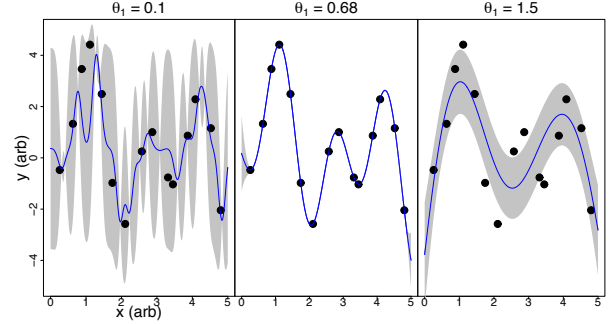


FIG. 3. Demonstration of emulator behavior as a function of correlation length,  $\theta_1$ . In all panels, the solid blue line shows the mean of the emulator and the solid gray region is a 95% confidence interval around this region. Left panel: fitting with a value of  $\theta_1$  that is too small (under-smoothing). Right panel: shows over-smoothing by using a value of  $\theta_1$  that is too large. Central panel: smoothing with a value of  $\theta_1 = 0.68$  that was obtained by a maximum likelihood estimation method.

the training values at those points. Note that the quantities defined in (26) depend implicitly upon the choice of correlation length scales  $\theta = \{\theta_0, \theta^k, \theta_N\}$  which determine the shape of the covariance function.

The expression for the mean in (26) can be decomposed into a contribution from the prior, the linear regression model  $\mathbf{h}(\mathbf{x})^T \hat{\beta}$  plus a contribution applied to the residuals determined by the covariance structure  $\mathbf{k}^T(\mathbf{x}) \mathbf{C}^{-1} (\mathbf{Y} - \mathbf{H} \hat{\beta})$ . Similarly the covariance can be decomposed into a contribution from the prior, the covariance function  $c(\mathbf{x}_i, \mathbf{x}_j)$  plus corrections arising from the prior covariance structure and the covariance of the new location  $\mathbf{x}$  through  $\mathbf{k}(\mathbf{x})$ . These terms weight the points  $\mathbf{x}_i, \mathbf{x}_j$  more highly the closer they are to the training points through  $\mathbf{k}$ . The  $\Gamma$  term gives the corrections to the covariance arising from the regression model.

In our study, we run the full code at  $N = 729$  (chosen because  $3^6 = 729$ ) points from the parameter space. A Latin Hyper Cube (LHC) design is used to generate the training locations in the parameter space. This is an efficient design for space-filling in high dimensional parameter spaces. This is an efficient sparse design for high dimensional parameter spaces that is “space-filling” in the sense that all its lower-dimensional projections are

distributed as evenly as possible [57–59].

The output from the model code is multivariate. Although fully multivariate emulator formulations do exist they are challenging to implement. Instead we follow the now somewhat standard procedure of creating emulators for some decomposition of the code output see eg [52, 60]. In this case we apply a principle components decomposition to the model output and build emulators for each significant component as detailed above.

## V. TESTING THE EMULATOR

The goal of this section is to investigate the reliability and accuracy of the Gaussian-process emulator described in the previous section. The emulators were tuned to full model calculations which were performed for 729 points in the six-dimensional parameter space. The points were chosen randomly according to a maximin Latin hypercube design [57]. The model output for each run was analyzed to determine the 15 summary observables,  $y_i$ , described in Section. III. The 15 numbers were transformed to variables  $\tilde{y}_i$ , which have zero mean and uncertainties of unity as described in (14). From these, six principal components  $z_i$  were then constructed as the linear combinations of the  $\tilde{y}$  with largest variance throughout the 729 sampling runs. The emulator’s task was to faithfully reproduce these six “observables” as a function of the six-dimensional parameter space  $\mathbf{x}$  along with providing a reasonable measure of uncertainty. To test the emulator, an additional 32 runs were performed using 32 random points in parameter space. A successful interpolation should reproduce the principal components from each of the 32 test runs to a given tolerance.

Tuning the Gaussian-process emulator involved estimating the hyper-parameters described in (24). The first attempt at finding optimized hyper-parameters used the same methods of [13, 56]. However, that approach was not robust, and often led to inaccurate emulators, with the accuracy defined by how well the emulator reproduced data from the 32 test runs. A more accurate result ensued by simply setting the hyper-radii, the  $\theta^i$  values in (24), equal to half the range for each parameter  $x^i$  in the model space. The exponent  $\alpha$  was set to 1.5 and the nugget  $\Theta_0$  was set to zero. Changing the hyper-radii by factors of two, or adjusting the exponent anywhere between 1.0 and 2.0 had little effect. For perspective, competing interpolating schemes were constructed, one based on a quadratic fit, and a second based on a linear fit where neighboring points were more heavily weighted in the fits. Each of these schemes was slightly less accurate than the Gaussian process emulator with the hyper-parameters chosen as described above. However, all these procedures performed better than the Gaussian process emulator using Maximum-Likelihood-Estimation (MLE) hyper-parameters as described in [13, 56]. This failure to find good hyper-parameters may come from the numerical challenges of the MLE optimization process given the

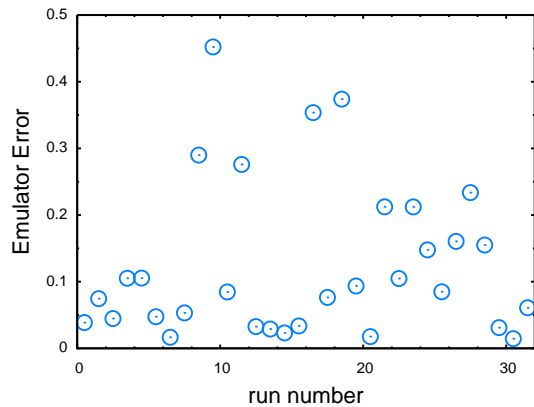


FIG. 4. (color online) The emulator error, EE, is shown for the 32 test runs. If the emulator error per principal component were of order of the experimental and model uncertainties, the values of EE would be near six. The errors above are significantly smaller.

large number of training points.

The Gaussian process emulator explicitly reproduces  $z_i(\mathbf{x})$  whenever  $\mathbf{x}$  approaches one of the training points,  $\mathbf{x}_n$ ,  $n = 1 \dots 729$ . To test the emulator points had to be chosen away from the training points, and 32 additional full model runs were performed at random points throughout the parameter space. The emulator error can be summarized as

$$EE(\mathbf{x}) \equiv \sum_{i=1}^r \left( z_i^{(\text{emu})}(\mathbf{x}) - z_i^{(\text{mod})}(\mathbf{x}) \right)^2, \quad (28)$$

where  $z_i^{(\text{emu})}(\mathbf{x})$  is the conditional mean from the  $i$ ’th emulator as given by (26) and  $r$  is the total number of observation principle components retained. A plot of this for the withheld data points is displayed in Figure. 4 for the 32 test runs. By construction,  $EE(\mathbf{x}_n)$  is zero for the 729 training runs. The fact that the net errors were less than unity, even after summing over six principal components, shows that the emulator did an outstanding job of reproducing the data. Furthermore, the emulator error is of the order of the statistical error of the model (which mainly comes from the calculation of  $v_2$ , which suggests that in this case further improving the emulator would not significantly improve the final result.

The success of the Gaussian Process emulation over a wide range of schemes and parameters is probably due to the smooth and monotonic response of the model to parameters. At high centre of mass energies, the physical system is highly explosive. Within the range of parameters considered the explosiveness is modified, but the behavior never changes qualitatively, and one expects a mostly linear response to the parameters. This may not be as true at lower energies.

It was seen that the estimate of the errors of the emulation as defined in (26) often significantly underestimated the accuracy of the emulator as tested in Figure. 4. The

net error tended to be less than a half unit, even though it was summed over multiple degrees of freedom. Since the error associated with the accuracy of the emulator was so small, the emulator error was incorporated into the calculation of the likelihood in a simplified manner. The uncertainty inherent to the data and models for a specific principal component was unity due to the choice in how to scale the  $z_i$  values. By adding in the emulator error, the total uncertainty should be  $\sigma^2 = 1 + \sigma_e^2$  for each component. The likelihood used by the MCMC is then,

$$\mathcal{L}(\mathbf{x}) \propto \exp \left\{ -\frac{1}{2} \sum_i \frac{(z_i^{(\text{emu})}(\mathbf{x}) - z_i^{(\text{exp})})^2}{1 + \sigma_e^2} \right\}. \quad (29)$$

For our MCMC calculations,  $\sigma_e$  was set to 0.1 according to an estimate of the error per degree of freedom from Fig. 4. This increased the width of the posterior region of parameter space by only a few percent.

## VI. MCMC RESULTS

As shown in the previous section, the emulator accurately reproduces the log-likelihood. For the MCMC search the Gaussian process emulator was run sampling many millions of points in parameter space. The trace provides an ergodic sample of the allowed regions in parameter space, i.e., the posterior distribution. The MCMC procedure applied here is a Metropolis algorithm. First, the parameter space was scaled and translated so that it was centered around zero, and that the flat prior had unit variance, i.e., it varied from  $-\sqrt{3}$  to  $+\sqrt{3}$ . First, a random point was chosen in the six-dimensional parameter space  $\mathbf{x}_1$ , from which one takes a random step to  $\mathbf{x}_2 = \mathbf{x}_1 + \delta\mathbf{x}$ . The random steps  $\delta\mathbf{x}$  were chosen according to a six-dimensional Gaussian with the step size in each dimension being 0.1. The likelihoods were calculated for each point. If the likelihood  $\mathcal{L}(\mathbf{x}_2)$  was higher than  $\mathcal{L}(\mathbf{x}_1)$ , the step was accepted, and if the likelihood was smaller, the step was accepted with the probability of the ratios of the two likelihoods. After the 100,000-step burn-in phase, the trace was stored by writing down every tenth point. The resulting distribution is proportional to the likelihood [61] and represents an ergodic sampling of the posterior distribution for a uniform prior. The trace finished when  $10^6$  points were written to disk.

To evaluate the success of the emulation, 20 points were randomly chosen from the MCMC trace and were then evaluated with the full model. The observables used for the original analysis were then plotted for each of the 20 points in parameter space. Another twenty points were chosen randomly from the original parameter space, i.e. they are consistent with the flat prior distribution. Again, the observables were calculated with the full model for each of these points in parameter space. One expects the observables for each of the 20 points representing the MCMC trace to reasonably well match

the experimental data, while the points chosen randomly from the prior distribution should lead to a wider range of observables, some of which should be inconsistent with the data.

Comparisons of the spectra from the model runs characterizing the prior and posterior distributions are shown in Figure. 5. Parameters from the posterior distributions lead to far superior fits, to both the yields and shape of the spectra. From the figure, one can see that the spectra for heavier particles provide more discriminating power. This comes from the greater sensitivity to collective flow, and emphasizes the importance of having reliable measurements of proton spectra. At RHIC, STAR's proton spectra are warmer than those of PHENIX, and their estimate of the mean  $p_t$  for protons is 7% higher. Whereas PHENIX shows the mean  $p_t$  of protons staying steady or perhaps slightly falling with increasing centrality, STAR's analysis show a rising mean  $p_t$ . If the mean  $p_t$  were indeed higher than what PHENIX reports, the extracted parameters should change, e.g., the initial collective flow might come out higher.

Figure 6 shows  $v_2$  and the femtoscopic source sizes calculated from the same representative points in parameter space for both the prior and posterior distributions. The MCMC is clearly successful in identifying points in parameter space that when run through the full model matched the experimental measurement. Further, given that the systematic uncertainty of specifying the  $p_t$  averaged  $v_2$  was assumed to be 12%, the spread of  $v_2$  vs.  $p_t$  plots appears consistent with expectations. Although the overall trend of the source radii were matched by the model, a consistent discrepancy between the data and model calculations using parameters from the posterior distribution is evident. At low  $p_t$ , the sideward and longitudinal source sizes are over-predicted by more than 10%, which is about double the expected systematic error. Over-predicting the sizes might be due to the assumption of boost invariance, or perhaps the approximations used to account for the Coulomb interaction, or in justifying the Koonin formula. From analyticity, one expects that the  $R_{\text{out}}$  and  $R_{\text{side}}$  sizes to approach one another as  $p_t \rightarrow 0$ . As can be seen in the upper panel of Figure. 6, this does not appear to be holding true in the data. Either the lower range of  $p_t$  (200 MeV/c) is not sufficiently small, or an acceptance/efficiency effect in the detector is affecting the result. This issue should be resolved if femtoscopic analyses are to be applied with confidence at better than the 10% level.

From the MCMC traces, the distribution of the various parameters and the correlations between pairs of parameters are shown for the Gaussian-process emulator in Figure. 7. The plots along the diagonal display the range of acceptable values for individual parameters, integrated over all values of the other five parameters. Although over 90% of the six-dimensional parameter space is eliminated at the one-sigma level, the individual parameters are rarely constrained to less than half their initial range when other parameters are allowed to vary.



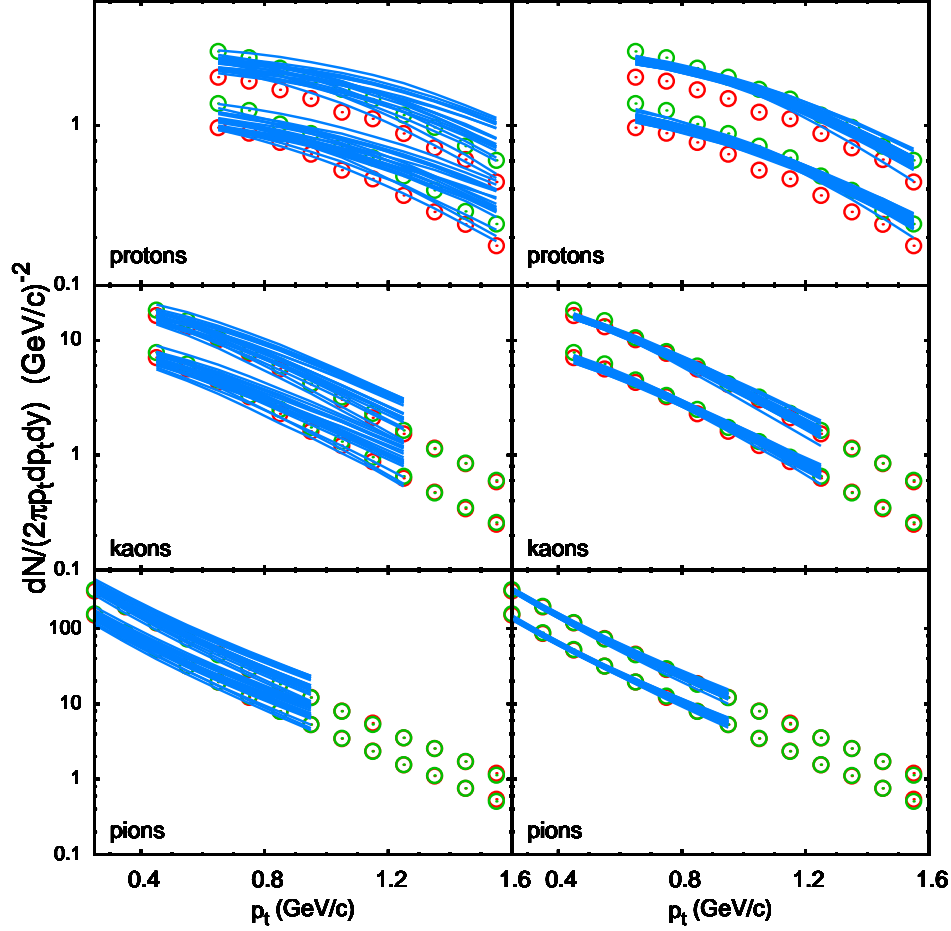


FIG. 5. (color online) Left-side panels: Pion, kaon and proton spectra from 20 model calculations where parameters are randomly chosen from the prior distribution. Model calculations are blue lines, and experimental data from PHENIX is shown as red/green circles for positive/negative charges. Results are shown for both 5% most central, and for the 20-30% centrality bin. Due to lack of some chemical reactions, normalizations for kaons and pions in the model were scaled by factors of 0.85 and 0.6 respectively. Right panels: Same as left-side but with 20 model calculations where parameters were chosen randomly from posterior distribution as sampled by MCMC trace.

The first four parameters (“I.C. PP NORM”, “I.C. SAT  $\sigma$ ”, “I.C. W.N. FRAC” and “I.C. FLOW”) determine the initial state fed into the hydro. The first parameter “I.C. PP NORM” sets the constant of proportionality between the product of the areal densities of the incoming nuclei, and the initial energy density fed into the hydro. In the limit of low aerial densities this should be consistent with  $pp$  collisions. Thus, the range of the prior distribution was quite small, and the statistical analysis did little to further constrain it. The parameter “I.C. SAT  $\sigma$ ” refers to  $\sigma_{\text{sat}}$  in (3) and parameterizes the saturation of the energy density with multiple collisions. The preferred value appears rather close to the value of 42 mb typically used in the wounded nucleon model, though there is a fairly wide range of accepted values. The parameter “I.C. W.N. FRAC” sets the weights between the wounded nucleon and the saturation parameterizations in (1). This shows a preference for the wounded nucleon prescription which gives a smaller initial anisotropy

than the saturation parameterization. The final initial-condition parameterization, “I.C. FLOW” sets the initial transverse flow set in the hydrodynamic calculation. The parameter sets the initial flow as a fraction of the amount described by (7), which should be expected in the limit of high-energy. The MCMC trace points to a rather small fraction of this flow, though like all of the initial-condition parameters has a fairly broad range of possible values.

The last two parameters refer to the viscosity. The viscosity at  $T = 170$  MeV is referred to as “ $\eta/s$ ” in Figure. 7, and the temperature dependence is labelled by “ $T$  DEP. of  $\eta$ ”, and refers to the parameter  $\alpha$  in (11). Both are significantly constrained as a fraction of the original parameter space. The range of  $\eta/s$  is consistent with similar, but less complete, searches through parameter space using similar models [2, 3]. In [62], the authors found little sensitivity to the viscosity at higher temperatures, but considered a smaller variation of the viscosity with temperature than was considered here.

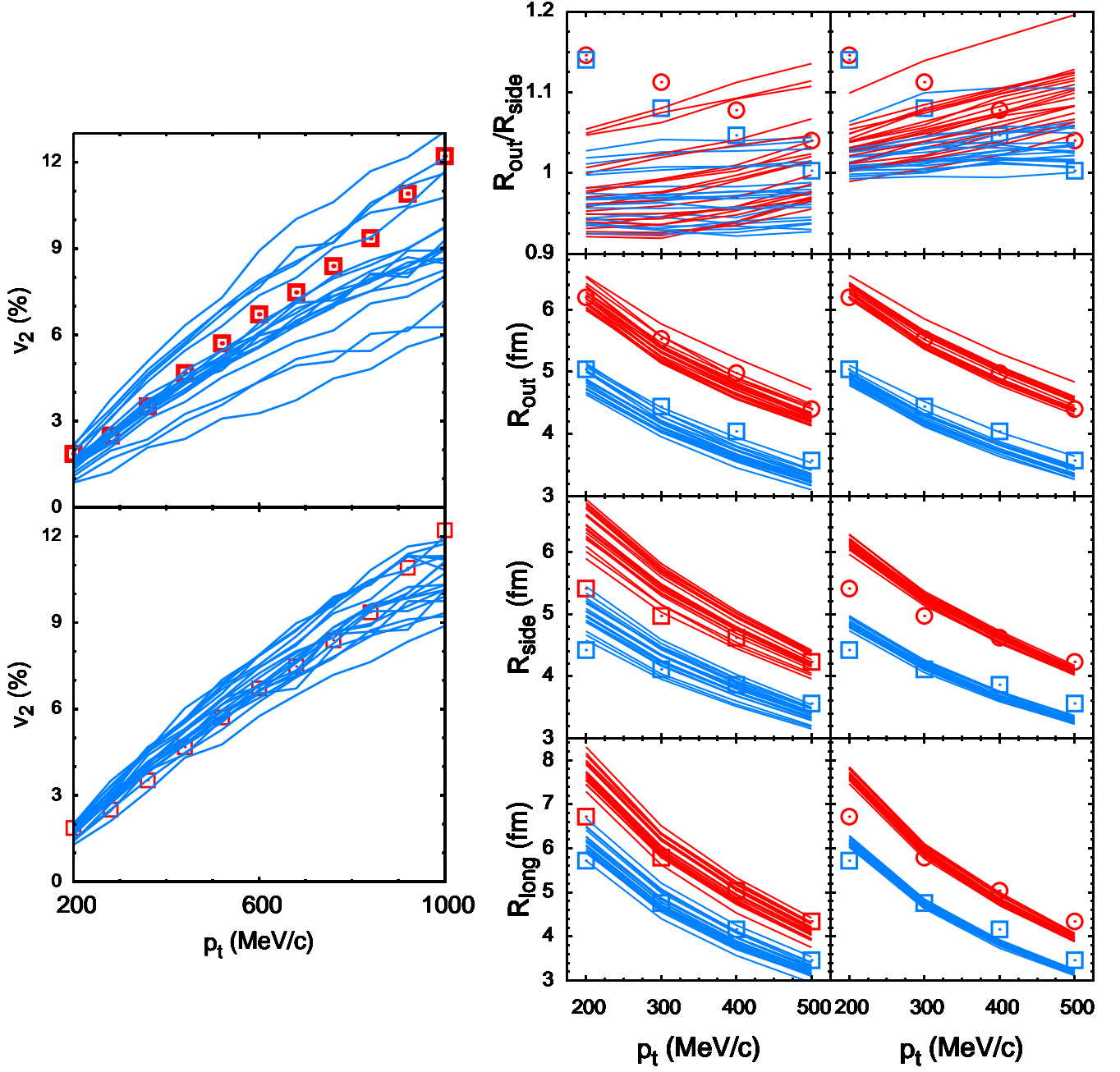


FIG. 6. (color online) Upper left panel: For 20 points in parameter space randomly chosen from the prior distribution,  $v_2$  for pions is plotted as a function of  $p_t$  for full model runs. Blue lines represent model calculations whereas are squares are experimental data. Lower left panel: Same as upper panel, except 20 points are randomly taken from posterior distribution as sampled by the MCMC trace. Right-side panels: Femtoscentic radii are shown for calculations from the prior distribution (left half) and from the posterior (right half) calculations. The posterior calculations well reproduce the data except for the sideward and longitudinal radii at low  $p_t$ .

Figure 7 also displays cross-correlations from the MCMC traces. Several parameters are strongly correlated. For instance, the energy normalization “I.C. PP NORM” and “I.C. SAT  $\sigma$ ” are strongly correlated in that one can have less saturation of the cross section if the energy normalization is turned down. There is also a strong correlation between “I.C. FLOW” and “I.C. W.N. FRAC”. One can compensate for less initial flow if the saturation formula is more heavily used than the wounded nucleon formula. Again, this is expected because the wounded nucleon parameterization leads to less spatial anisotropy and a somewhat more diffuse initial state.

The inferred viscosity is clearly correlated with the weighting between the wounded nucleon and saturation parameterizations, as expected from the arguments in [26]. The two viscous parameters are also correlated with one another as expected. One can compensate for a very low viscosity at  $T = 170$  MeV by having the viscosity rise quickly with temperature. Figure 8 shows the viscosity to entropy ratio as a function of temperature corresponding to the 20 random samples from the both the prior and posterior distributions. Higher values of the temperature dependence  $\alpha$  are increasingly unlikely for higher values of  $\eta/s|_{T_c}$ .

## VII. SUMMARY AND OUTLOOK

Two principal conclusions can be taken from this study. First, the data from relativistic heavy ion collisions is well suited to a multi-dimensional analysis featuring model emulators. The response of the data to model parameters appears sufficiently smooth to warrant simple interpolation of a few principle components. Only a half dozen parameters were varied in this study, and only a limited number of observables were considered. Nonetheless, the procedure should easily scale to larger numbers of parameters and larger data sets. The success of the emulators in reproducing model output, and in the MCMC procedure in identifying likely regions in parameter space provide hope that the field can produce quantitative statements concerning the bulk properties and dynamics of the matter formed in heavy ion collisions. The second conclusion centers on the extracted parameters. Although the ranges are subject to change given expected improvements in both data and modeling, the ranges of parameters and correlations shown in Figure. 7 are remarkably close to expectations from less rigorous searches.

The statistical procedures applied here represent a significant improvement to the state-of-the-art for comparisons of data and models in the field of relativistic heavy ion physics. Previously, parameters were varied either individually, or in small groups. Figure 4 demonstrates the success of using emulators for this problem. Most importantly, the emulator techniques should scale well with increased data and increased number of param-

eters. Ultimately, the number of simultaneously varied parameters might increase to around 20, with the expectation that many, or perhaps most, of these parameters will not be significantly constrained by the data. Additional parameters to which the model is insensitive does not increase the need for additional runs, as long as the parameters are varied in such a way that those parameters that are important are well sampled. It is expected that the additional parameters would decrease the efficiency with which the critical parameters are sampled, but that the additional number of model runs would not make the problem intractable. Adding more data should, hopefully, increase the number of principal components extracted from the model runs by providing additional discriminating power. Once the model calculations have been performed, the numerical cost of the statistical analysis performed here was negligible, and adding more principal components should not cause any problems. Thus, the results of this study are promising, and encourage extending the scope to much larger data sets and more realistic models.

Although the models used here represented the current state of the art, several improvements are necessary before firm quantitative conclusions can be extracted. The following improvements require significant development, but are all tractable.

- A flexible equation of state. For this study, the equation of state was fixed. There is some uncertainty involved with lattice calculations that should be accounted for with a variable equation of state. Additionally, it is of interest to address whether the equation of state is constrained by experiment alone, i.e. without relying on lattice calculations.
- The bulk viscosity was set to zero here. Near  $T_c$ , the system undergoes a rapid change in microscopic structure, and the system may lose equilibrium. As the system returns to equilibrium, entropy is generated. If the departure from equilibrium is small, the effect can be accounted for by adding a bulk viscosity [17, 18]. If the departure is large, other approaches are possible, such as dynamically solving for the mean fields [63].
- The chemical composition of the hadronic phase was set by the assumption of chemical equilibrium when the system reached a temperature of 170 MeV. The chemical evolution can be improved by incorporating more inelastic process like baryon-antibaryon annihilation [64, 65]. Further, the assumption of perfect equilibrium at a fixed temperature should be relaxed by parameterizing non-equilibrium effects.
- Although collisions produce many thousands of particles, the initial collisions involves only on the order of 100 nucleons. The finite number of original scattering centers leads to lumpy initial conditions, unlike the smooth initial conditions used here. If the model used here were improved to incorporate



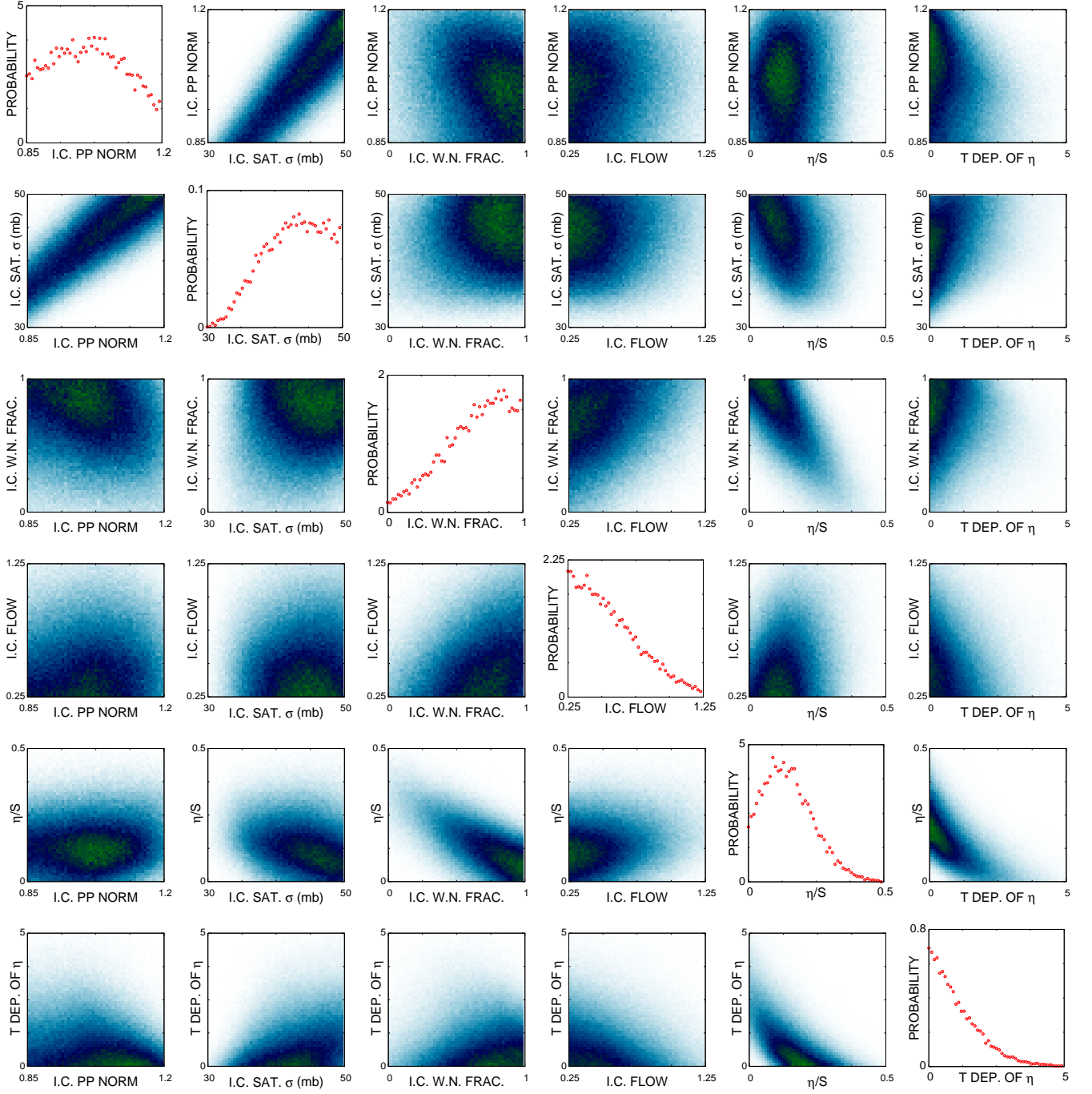


FIG. 7. (color online) The distribution of acceptable values for each of the six model parameters are shown along the diagonal. The off-diagonal plots display the correlation between all pairs of observables. Four of the six parameters refer to the initial state (hence the “I.C.” in their name) and the last two describe the shear viscosity.

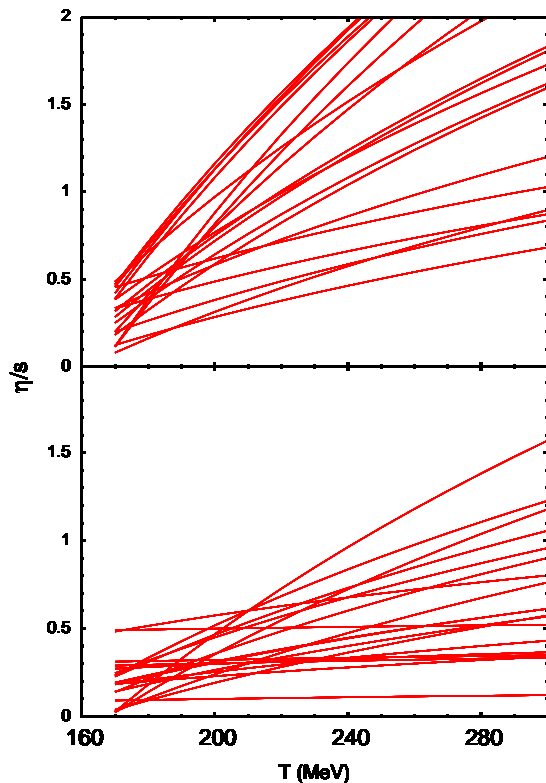


FIG. 8. (color online) Twenty random points in parameter space were taken from the prior (upper panel) and posterior (lower panel) distributions. The temperature dependence of the viscosity to entropy ratios is clearly constrained by the statistical comparison with data, though the posterior distribution still covers a large variation.

initial fluctuations,  $v_2$  would be more realistically modeled, and it would make it possible to consider fluctuations of the flow encoded in higher harmonics, i.e.,  $v_3, v_4 \dots$  [23].

- Although the model used here can incorporate three-dimensional flow, for this study the calculations were performed with the Bjorken ansatz. This approximation is reasonable for collisions at 200.4 GeV or higher [35], but full three-dimensional calculations are needed for lower energy, or for observables away from mid-rapidity. Using low-energy or non-mid-rapidity data will also necessitate a more complex parameterization of the initial state.
- During the hadronic phase, pionic phase space becomes highly filled at low  $p_t$ . This affects spectra at the 10% level, which is neither crucial or non-negligible. Hadronic cascades can incorporate such effects by adding  $(1+f)$  phase space enhancements to scatterings. This increases the numerical cost of the modeling at the factor of two level.

Most of the improvements listed above would be accompanied by an increase in the number of parameters. Varying the equation of state will involve the addition of a few parameters. Since the bulk viscosity is not well determined by lattice calculations, both it and its temperature dependence require parameterization. Non-equilibrium chemistry can be parameterized by adding fugacities for the initial state. For instance, the initial conditions away from mid-rapidity, or for collisions at lower energy, could necessitate a half dozen new parameters. At lower energy, the dependence of the equation of state on baryon number is unknown and requires parameterization. The initial conditions at the LHC requires additional parameters to encapsulate the beam energy dependence of the initial density and flow profiles. The lumpiness of the initial state involves setting a transverse size of the fluctuations. It is easy to imagine future analyses involving on the order of 20 parameters.

The amount of available data for analyses such as these has swelled in the past few years. The beam-energy scan at RHIC provides all the observables analyzed here at a half dozen more energies. Additionally, Cu+Cu, Cu+Au and U+U collisions have been measured at RHIC. Additionally, results from Pb+Pb collisions at the LHC have now been analyzed and published. Finally, higher flow harmonics,  $v_2, \dots, v_n$ , have also become available.

Expanding the scope of the analysis to a larger range of beam energies and to include initial state fluctuations could increase the numerical cost of the calculations by two orders of magnitude. In the present study one processing core could perform a full model run for one point in parameter space in approximately one day. This would increase to being on the order of several weeks, or one month, if the beam energy scan, LHC data, and initial-state fluctuations were included. If the number of points sampled in parameter space were increased to a few thousand to better account for the larger number of parameters, the project would remain tractable, but would clearly require significant allocation of resources. The success and scalability of the methods presented here suggest that such an effort could help transform heavy ion physics into a more rigorously quantitative science.

## ACKNOWLEDGMENTS

This work was supported by the National Science Foundation's Cyber-Enabled Discovery and Innovation Program through grant NSF-0941373 and by the Department of Energy Office of Science through grant number DE-FG02-03ER41259. RW received additional support from the National Science Foundation Division of Mathematical Sciences through grant NSF-DMS-1228317 and from the National Aeronautics and Space Administration, grant no. NNX09AK60G

- 
- [1] P. Romatschke and U. Romatschke, Phys. Rev. Lett. **99**, 172301 (2007) [arXiv:0706.1522 [nucl-th]].
- [2] R. A. Soltz, I. Garishvili, M. Cheng, B. Abelev, A. Glenn, J. Newby, L. A. L. Levy and S. Pratt, to appear in Physical Review C, arXiv:1208.0897 [nucl-th].
- [3] U. W. Heinz, C. Shen and H. Song, AIP Conf. Proc. **1441**, 766 (2012) [arXiv:1108.5323 [nucl-th]].
- [4] C. Shen, S. A. Bass, T. Hirano, P. Huovinen, Z. Qiu, H. Song and U. Heinz, J. Phys. G **38**, 124045 (2011) [arXiv:1106.6350 [nucl-th]].
- [5] C. Shen, U. Heinz, P. Huovinen and H. Song, Phys. Rev. C **84**, 044903 (2011) [arXiv:1105.3226 [nucl-th]].
- [6] H. Song, S. A. Bass and U. Heinz, Phys. Rev. C **83**, 054912 (2011) [arXiv:1103.2380 [nucl-th]].
- [7] H. Song, S. A. Bass and U. Heinz, Phys. Rev. C **83**, 024912 (2011) [arXiv:1012.0555 [nucl-th]].
- [8] H. Song, S. A. Bass, U. Heinz, T. Hirano and C. Shen, Phys. Rev. Lett. **106**, 192301 (2011) [arXiv:1011.2783 [nucl-th]].
- [9] C. Shen, U. Heinz, P. Huovinen and H. Song, Phys. Rev. C **82**, 054904 (2010) [arXiv:1010.1856 [nucl-th]].
- [10] H. Song and U. W. Heinz, Phys. Rev. C **81**, 024905 (2010) [arXiv:0909.1549 [nucl-th]].
- [11] M. D. Schneider, L. Knox, S. Habib, K. Heitmann, D. Higdon and C. Nakhleh, Phys. Rev. D **78**, 063529 (2008) [arXiv:0806.1487 [astro-ph]].
- [12] S. Habib, K. Heitmann, D. Higdon, C. Nakhleh and B. Williams, Phys. Rev. D **76**, 083503 (2007) [astro-ph/0702348 [ASTRO-PH]].
- [13] F. A. Gomez, C. E. Coleman-Smith, B. W. O'Shea, J. Tumlinson and R. L. Wolpert, arXiv:1209.2142 [astro-ph.GA].
- [14] H. Petersen, C. E. Coleman-Smith, S. A. Bass and R. Wolpert, J. Phys. G **38**, 045102 (2011) [arXiv:1012.4629 [nucl-th]].
- [15] S. Borsanyi, G. Endrodi, Z. Fodor, A. Jakovac, S. D. Katz, S. Krieg, C. Ratti and K. K. Szabo, JHEP **1011**, 077 (2010) [arXiv:1007.2580 [hep-lat]].
- [16] A. Bazavov, T. Bhattacharya, M. Cheng, N. H. Christ, C. DeTar, S. Ejiri, S. Gottlieb and R. Gupta *et al.*, Phys. Rev. D **80**, 014504 (2009) [arXiv:0903.4379 [hep-lat]].
- [17] K. Paech and S. Pratt, Phys. Rev. C **74**, 014901 (2006) [nucl-th/0604008].
- [18] F. Karsch, D. Kharzeev and K. Tuchin, Phys. Lett. B **663**, 217 (2008) [arXiv:0711.0914 [hep-ph]].
- [19] L. D. McLerran and R. Venugopalan, Phys. Rev. D **49**, 2233 (1994) [hep-ph/9309289].
- [20] D. Kharzeev and M. Nardi, Phys. Lett. B **507**, 121 (2001) [nucl-th/0012025].
- [21] D. Kharzeev, E. Levin and M. Nardi, Nucl. Phys. A **730**, 448 (2004) [Erratum-ibid. A **743**, 329 (2004)] [hep-ph/0212316].
- [22] A. Dumitru, E. Molnar and Y. Nara, Phys. Rev. C **76**, 024910 (2007) [arXiv:0706.2203 [nucl-th]].
- [23] B. Schenke, P. Tribedy and R. Venugopalan, Phys. Rev. Lett. **108**, 252301 (2012) [arXiv:1202.6646 [nucl-th]].
- [24] For an overview of wounded nucleon models see M. L. Miller, K. Reygers, S. J. Sanders and P. Steinberg, Ann. Rev. Nucl. Part. Sci. **57**, 205 (2007) [nucl-ex/0701025].
- [25] B. I. Abelev *et al.* [STAR Collaboration], Phys. Rev. C **79**, 034909 (2009) [arXiv:0808.2041 [nucl-ex]].
- [26] H. -J. Drescher, A. Dumitru, C. Gombeaud and J. -Y. Ollitrault, Phys. Rev. C **76**, 024905 (2007) [arXiv:0704.3553 [nucl-th]].
- [27] S. Pratt, Phys. Rev. Lett. **102**, 232301 (2009) [arXiv:0811.3363 [nucl-th]].
- [28] J. Vredevoogd, Ph.D. Thesis, Michigan State University, in preparation (2013).
- [29] J. Vredevoogd and S. Pratt, Phys. Rev. C **79**, 044915 (2009) [arXiv:0810.4325 [nucl-th]].
- [30] J. Vredevoogd and S. Pratt, Nucl. Phys. A **830**, 515C (2009) [arXiv:0907.4733 [nucl-th]].
- [31] A. Muronga, Phys. Rev. Lett. **88**, 062302 (2002) [Erratum-ibid. **89**, 159901 (2002)] [arXiv:nucl-th/0104064].
- [32] K. Dusling and D. Teaney, Phys. Rev. C **77**, 034905 (2008) [arXiv:0710.5932 [nucl-th]].
- [33] H. Song and U. W. Heinz, Phys. Rev. C **77**, 064901 (2008) [arXiv:0712.3715 [nucl-th]].
- [34] M. Luzum and P. Romatschke, Phys. Rev. C **78**, 034915 (2008) [Erratum-ibid. C **79**, 039903 (2009)] [arXiv:0804.4015 [nucl-th]].
- [35] J. Vredevoogd and S. Pratt, Phys. Rev. C **85**, 044908 (2012) [arXiv:1202.1509 [nucl-th]].
- [36] S. Pratt and G. Torrieri, Phys. Rev. C **82**, 044901 (2010) [arXiv:1003.0413 [nucl-th]].
- [37] S.A.Bass et al [URQMD], Progr. Part. Nucl. Physics Vol. **41**, 225 (1998).
- [38] Boltzmann 3D, the B3D code is available if you email Scott Pratt, prattsc@msu.edu.
- [39] Particle Data Group, *The Review of Particle Physics*, <http://pdg.lbl.gov> (2012).
- [40] S. E. Koonin, Phys. Lett. B **70**, 43 (1977).
- [41] M. A. Lisa, S. Pratt, R. Soltz and U. Wiedemann, Ann. Rev. Nucl. Part. Sci. **55**, 357 (2005) [nucl-ex/0505014].
- [42] Correlations Analysis Library, Scott Pratt and David Brown. The software is available by contacting Scott Pratt, prattsc@msu.edu.
- [43] B. Schenke, S. Jeon and C. Gale, Phys. Rev. Lett. **106**, 042301 (2011) [arXiv:1009.3244 [hep-ph]].
- [44] B. Schenke, S. Jeon and C. Gale, arXiv:1109.6289 [hep-ph].
- [45] J. Adams *et al.* [STAR Collaboration], Phys. Rev. C **72**, 014904 (2005) [nucl-ex/0409033].
- [46] B. I. Abelev *et al.* [STAR Collaboration], Phys. Rev. C **80**, 024905 (2009) [arXiv:0903.1296 [nucl-ex]].
- [47] S. S. Adler *et al.* [PHENIX Collaboration], Phys. Rev. C **69**, 034909 (2004) [nucl-ex/0307022].
- [48] A. O'Hagan, Reliability Engr. & System Safety **91**, 1290 (2006).
- [49] J.E. Oakley and A. O'Hagan, Biometrika **89**, 769 (2002).
- [50] J.E. Oakley and A. O'Hagan, J. of the Royal Stat. Soc., **B66**, 751 (2004).
- [51] M.C. Kennedy and A. O'Hagan, Biometrika **87**, 1 (2000).
- [52] D. Higdon, J. Gattiker, B. Williams and M. Rightley, J. of the Amer. Stat. Assoc. **103**, 570 (2008).
- [53] M.J. Bayarri *et al.*, "A Framework for Validation of Computer Models", Technical Report (2002).
- [54] J.-P. Chilès and P. Delfiner, "Geostatistics: Modeling Spatial Uncertainty", John Wiley & Sons, New York, NY (1999).

- [55] N. Cressie, “Statistics for Spatial Data”, John Wiley & Sons, New York, NY (1993).
- [56] C.E. Rasmussen and C.K.I. Williams, Gaussian Processes for Machine Learning, The MIT Press (2005).
- [57] J. Sacks, W.J. Welch, T.J. Mitchell and H.P. Wynn, *Stat. Sci.* **4**, 409 (1989).
- [58] T.J. Santner, B.J. Williams and W. Notz, “The Design and Analysis of Computer Experiments”, Springer Verlag, NY (2003).
- [59] McKay, M.D.; Beckman, R.J.; Conover, W.J., *Technometrics* 21(2) 239–245. DOI: 10.2307/1268522 (1979).
- [60] M.J. Bayarri et al., “Computer Model Validation with Functional Output”, Technical Report (2006).
- [61] S.E. Koonin: Computational Physics, Benjamin/Cummings (Menlo Park, CA) 1986.
- [62] H. Niemi, G. S. Denicol, P. Huovinen, E. Molnar and D. H. Rischke, *Phys. Rev. Lett.* **106**, 212302 (2011) [arXiv:1101.2442 [nucl-th]].
- [63] K. Paech, H. Stoecker and A. Dumitru, *Phys. Rev. C* **68**, 044907 (2003) [nucl-th/0302013].
- [64] Y. Pan and S. Pratt, arXiv:1210.1577 [nucl-th].
- [65] J. Steinheimer, J. Aichelin and M. Bleicher, arXiv:1203.5302 [nucl-th].